International Journal of Advanced Engineering Research and Science (IJAERS)
Peer-Reviewed Journal
ISSN: 2349-6495(P) | 2456-1908(O)
Vol-9, Issue-12; Dec, 2022
Journal Home Page Available: https://ijaers.com/
Article DOI: https://dx.doi.org/10.22161/ijaers.912.1

Analysis of content published on the social network Twitter related to Brazilian candidates for the presidency of the republic

Análise de conteúdos publicados na rede social Twitter relacionados aos candidatos brasileiros à presidência da república

Renato Lopo Montalvão Júnior, Victória Felipe Damascena Souza, Raphael Magalhães Hoed

Instituto Federal de Educação, Ciência e Tecnologia do Norte de Minas Gerais – Campus Januária

Received: 04 Nov 2022,

Receive in revised form: 27 Nov 2022,

Accepted: 04 Dec 2022,

Available online: 09 Dec 2022

***Keywords*— Sentiment analysis, Elections, Data mining, Social network.**

***Palavras-chave*— Análise de sentimento, Eleições, Mineração de dados, Rede Social.**

***Abstract*— *This article seeks to verify how related the opinion of Twitter users about the Brazilian candidates for the presidency of the republic is with the polls of voting intentions. Do the positive and negative sentiments on Twitter track the results of the polls of voting intentions for president? In view of this, publications made on Twitter that mention candidates for the presidency of the republic were collected, using python scripts to collect the tweets and to classify them supervised machine learning algorithms via the sklearn library. Therefore, it is intended to verify how compatible the interactions on Twitter are with the results of the presidential polls of the year 2022.*

***Resumo*— *Este artigo, busca verificar o quão relacionada está a opinião dos usuários do Twitter sobre os candidatos brasileiros à presidência da república com as pesquisas de intenção de voto. Os sentimentos positivos e negativos no Twitter acompanham os resultados das pesquisas de intenção de voto para presidente? Diante disso, foram coletadas publicações feitas no Twitter que citam os candidatos à presidência da república, usando para a coleta dos tweets scripts python e para classificá-los algoritmos de aprendizagem de máquina supervisionado via biblioteca sklearn. Portanto, pretende-se verificar o quão compatível as interações no Twitter são com os resultados das pesquisas presidenciais do ano de 2022.*

I. INTRODUÇÃO

Os candidatos à Presidência da república ou candidatos políticos no geral sempre buscam meios de se promoverem e venderem uma boa imagem, seja através de propagandas de rádio, TV (Televisão), panfletos ou até mesmo com passeatas na rua. Nos dias atuais, com advento da Internet e das redes sociais, os candidatos encontram um novo meio de se promoverem, usando as redes sociais para passar uma boa imagem, se posicionar sobre acontecimentos atuais ou até mesmo atacar a oposição. A

rede social Twitter atualmente é uma das maiores redes sociais em questão de usuários ativos, e também é um local que grande parte dos usuários usam para se expressar politicamente. O Twitter é uma rede social on-line caracterizada por permitir a postagem apenas de textos curtos, 280 caracteres no máximo.

Diante desse cenário, esse trabalho busca medir se a opinião dos usuários do Twitter sobre os candidatos à presidência da república se comporta de maneira semelhante às pesquisas de intenção de voto. Os sentimentos positivos e negativos no Twitter acompanham os resultados das pesquisas de intenção de voto para presidente? O engajamento do perfil do candidato com o público segue o mesmo padrão que as pesquisas eleitorais? Nesta investigação, o foco é coletar publicações feitas no Twitter relacionadas aos 4 candidatos com maior percentual nas pesquisas de intenção de voto divulgadas pelos principais institutos de pesquisa, sendo eles: Jair Bolsonaro (PL - Partido Liberal), Lula (PT - Partido dos Trabalhadores), Ciro Gomes (PDT - Partido Democrático Trabalhista) e Simone Tebet (MDB - Movimento Democrático Brasileiro), onde serão coletados tweets em que o candidato é mencionado.

Na Tabela 1 são apresentados os nomes dos candidatos e a quantidade de seguidores na rede social Twitter que o candidato possuía até o dia 18 de setembro de 2022.

Tabela 1. Candidatos à presidência, perfis e seguidores

Candidato	*User* no *Twitter*	Número de seguidores
Jair Bolsonaro (PL)	@jairbolsonaro	8.9 Milhões
Lula (PT)	@LulaOficial	4.3 Milhões
Ciro Gomes (PDT)	@cirogomes	1.4 Milhões
Simone Tebet (MDB)	@simonetebetbr	366.2 Mil

Serão usados como fontes de pesquisa de intenção de voto os institutos de análise estatística Datafolha[1], PoderData[2] e Ipespe[3]. Os critérios de escolha dos respectivos institutos é detalhado na Metodologia. Com a pesquisa proposta será possível observar como o sentimento por parte dos usuários do *Twitter* se manifesta durante o período de campanha eleitoral e à medida em que os resultados das pesquisas de intenção de votos vão sendo divulgados.

O objetivo geral deste trabalho é desenvolver uma pesquisa acerca do sentimento dos usuários da rede social Twitter sobre os candidatos à presidência da república e o quão estes sentimentos se assemelham com as pesquisas de intenção de voto conduzidas pelos principais institutos de pesquisa.

Os objetivos específicos consistem em:

- Criar um processo automatizado para coletar publicações feitas no Twitter, relacionadas aos candidatos listados na Tabela 1.
- Identificar sentimentos associados aos conteúdos coletados.
- Analisar e caracterizar conteúdos publicados no Twitter relacionados aos candidatos.

II. REFERENCIAL TEÓRICO

Essa etapa, tem como propósito, abordar conceitos que tratam do assunto ou que dêem embasamento teórico para o desenvolvimento, que serão tratados em quatro subseções: A rede social Twitter; biblioteca para coletar dados no Twitter; análise de sentimentos e mineração de dados.

2.1 A REDE SOCIAL TWITTER

Uma rede socal on-line pode ser definida como um sistema que proporciona um ambiente na Internet para que usuários estabeleçam relação e comunicação com outros [3]. O Twitter é uma rede social on-line, de serviço microblog que permite aos seus usuários enviarem e receberem mensagens de texto, chamadas "tweets". É possível enviar tweets através do site do Twitter ou através de uma aplicação móvel. Os tweets podem conter até 280 caracteres e são visíveis para todos os outros usuários do Twitter. Essa é uma das redes sociais mais populares que foi criada em 2006 por Jack Dorsey, Evan Williams, Biz Stone e Noah Glass, porém, só foi lançada publicamente em julho de 2007.

A rede social Twitter é considerada uma das principais ferramentas de comunicação do mundo. Em um ambiente no qual a informação é cada vez mais rápida e acessível, o Twitter se tornou uma plataforma de grande importância para a divulgação de notícias, opiniões e conteúdos.

[1] https://datafolha.folha.uol.com.br/
[2] https://www.poder360.com.br/
[3] https://ipespe.org.br/

O Twitter está se tornando cada vez mais um lugar para a política. Muitas pessoas usam o Twitter, como meio de seguir as últimas notícias, participar de "debates" e compartilhar pontos de vista.

Segundo Delcia Mattos Vidal [11], essa ferramenta tem sido utilizada para conquistar a opinião pública e fortalecer a imagem dos políticos frente aos eleitores, proporcionando um espaço para figuras públicas se aproximarem de seu público online. A plataforma serve como um palco para os políticos promoverem suas propostas e ações, incentivar debates e também receber críticas.

Essa mídia social foi uma parte importante da bem-sucedida campanha eleitoral de Barack Obama à presidência dos Estados Unidos, e desde então, boa parte dos políticos voltaram suas atenções ao uso das novas tecnologias [1].

Grandes líderes e autoridades usam Twitter de diversas maneiras. Alguns usam a rede para anunciar decisões oficiais, outros para se comunicar diretamente com os cidadãos e ainda há os que usam para compartilhar suas agendas e opiniões sobre os assuntos do momento. Donald Trump, ex-presidente dos Estados Unidos da América, era um usuário contumaz do Twitter e usava a rede para compartilhar suas opiniões sobre assuntos que estavam em alta. Já o atual presidente do Brasil, Jair Bolsonaro, usa o Twitter para se comunicar diretamente com os cidadãos, anunciar decisões oficiais e também compartilhar sua opinião sobre variados assuntos.

A seguir são relacionados os termos utilizados pelo Twitter que definem o funcionamento dessa rede social [4]:

- Perfil: conjunto de atributos (nome, descrição, foto e localização) declarados por um usuário no Twitter, sendo que todo perfil de usuário é identificado por um nome iniciado pelo caractere '@'.
- Tweet: é o nome dado a uma postagem publicada por um usuário.
- Seguindo (following): quando o usuário A está seguindo o usuário B, implica que A acompanha os tweets de B.
- Seguidor (follower): quando o usuário C é seguidor do usuário D, implica que que C acompanha os tweets de D.
- Timeline: é o conjunto de tweets que um usuário publicou no Twitter.
- Retweet: é um tweet publicado por um usuário A que foi compartilhado por outro usuário B.
- Hashtag: é um dos recursos mais populares do Twitter para rotular ou resumir um tema de discussão, são expressões iniciadas com o caractere '#'.

2.2 BIBLIOTECA PARA COLETAR DADOS NO TWITTER

Biblioteca é um acervo de itens que pode emprestar a seus usuários qualquer tipo de material, desde filmes e livros até músicas e jogos. Uma biblioteca para coletar dados do Twitter poderia incluir um software especializado para baixar tweets, uma API (application programming interface) do Twitter por meio da qual você pode programaticamente buscar pelos tweets mais recentes ou uma interface gráfica baseada na web que permita navegar e pesquisar pelos tweets armazenados em uma determinada biblioteca.

O Sampled stream v1 é um recurso da API oficial do Twitter que permite coletar tweets em tempo real, ou seja, no momento em que o tweet foi postado. É possível coletar cerca de 1% de todos os novos tweets ao enviar requisição do tipo GET para a URL https://api.twitter.com/labs/1/tweets/stream/sample. Conforme especificação da biblioteca é possível fazer até 50 requisições para coletar tweets em um intervalo de 15 minutos. A API permite fazer filtros ao requisitar coleta de dados, alguns exemplos: palavras-chave, hashtag, localização e tamanho do tweet[4].

2.3 ANÁLISE DE SENTIMENTOS

A análise de sentimentos envolve descobrir a emoção expressa no texto. As ferramentas de análise de sentimentos processam o texto e o classificam como contendo sentimento positivo, negativo ou neutro. Para fazer isso, essas ferramentas utilizam um dicionário de palavras com pesos de sentimento associados a elas [9].

Esta técnica é útil para pesquisas de opinião de grande escala, pois é não invasiva, rápida, autêntica, barata e automática. Uma forma de realizar pesquisa de opinião dessa maneira é a aplicação da análise de sentimento em opiniões ou comentários de produtos, serviços, eventos e marcas em mensagens de redes sociais [10].

Em relação aos estudos correlatos sobre análise de sentimento no Twitter, foi apresentado por [2] uma comparação entre 8 métodos que são bastante utilizados, sendo eles: SentiWordNet, SASA, PANAS-t, Emoticons, SentiStrength, LIWC, SenticNet, e Happiness Index. Nesse

[4] https://developer.twitter.com/en/docs/labs/sampled-stream/a pi-reference/get-tweets-stream-sample

estudo de comparação foi utilizado conteúdos da Web para detectar a polaridade (positivo e negativo). Foi concluído que os 8 métodos têm graus de abrangência e acurácia variados, e que não existe um método com melhores resultados. Além disso, foi construído um novo método chamado "Método Combinado", em que consiste na combinação dos métodos para alcançar uma acurácia maior.

Com a pandemia do COVID-19 houve um aumento nas buscas sobre o tema na internet. E as redes sociais, especialmente o Twitter, possibilitam o acesso e compartilhamento de conteúdo relacionados à pandemia, mas também acabam sendo um meio de propagação de notícias falsas. Diante desse contexto [7] analisa o sentimento dos usuários do Twitter em relação a pandemia da Covid-19. Utilizando processamento de linguagem natural para fazer análise de sentimento e coletando tweets no período de janeiro a junho de 2020, foi demonstrado como resultado que o Twitter tem um potencial como ferramenta de vigilância de surtos e epidemias, revelando que existe uma relação entre publicações do Twitter e o movimento de disseminação do coronavírus no Brasil. Denotando a preocupação dos usuários do Twitter em relação ao avanço da doença e as consequências por ela gerada, o sentimento negativo foi o mais dominante.

Segundo [8] textos em português tem uma certa dificuldade na análise de sentimentos por ter uma complexidade gramatical e ser rica em vocábulos. E os autores apresentam um comparativo entre diversos algoritmos de aprendizagem de máquina e de abordagem léxica, classificando tweets em português associados às Olimpíadas de 2016. Como resultado, foi mostrado que a aprendizagem de máquina se sobressai melhor que abordagem léxica, com o algoritmo SVM (Support Vector Machine) atingindo o melhor desempenho de 89,5% de acurácia no melhor cenário.

2.4 MINERAÇÃO DE DADOS

A mineração de dados (data mining) e a descoberta do conhecimento em banco de dados (knowledge discovery in database - KDD), constituem disciplina que define métodos para a preparação, análise e compreensão dos dados minerados [5]. [12] definem a mineração de dados como a procura de informações relevantes em enormes bancos de dados, em que se implica ato correlato entre o homem e máquina, onde a parte humana é planejar e criar os banco de dados, mensurar os problemas e definir as metas; já a máquina, apura os dados e buscam por padrões que entrem em conformidade com as os objetivos estipulados. A descoberta de conhecimento em bases de dados está diretamente ligada à procura de relações implícitas presentes nos dados verificados.

A mineração de dados é uma técnica de análise de dados que tem por objetivo extrair informações úteis de um conjunto de dados. É uma técnica estatística que usa algoritmos para explorar padrões em conjuntos de dados [6].

Essas técnicas de mineração de dados são divididas em três grandes grupos:

- Classificação: usada para predizer a categoria de um determinado grupo de dados;
- Associação: usada para agrupar dados em classes;
- Agrupamento: usada para agrupar dados em classes.

Neste trabalho será usado a técnica de classificação, pois o objetivo é classificar o tweet como o sentimento de negativo ou positivo.

III. METODOLOGIA

Neste trabalho foram empregados procedimentos metodológicos de pesquisa experimental para atingir os objetivos anteriormente descritos. É um trabalho classificado como de natureza aplicada no qual se empreenderá práticas que seguem uma abordagem quantitativa.

As etapas da metodologia considerada neste trabalho são apresentadas no fluxograma da Figura 1 e detalhadas na próxima seção (Etapas da Metodologia):
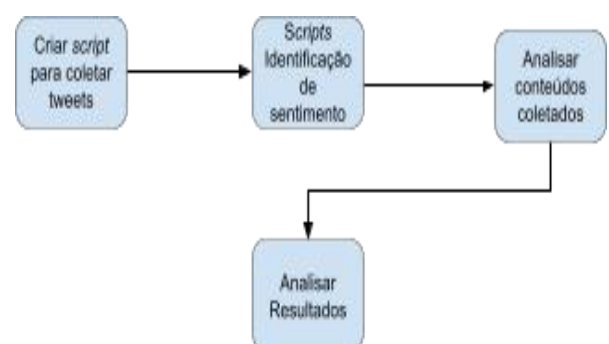
Fig 1. Etapas da Metodologia

3.1 ETAPAS DA METODOLOGIA

A metodologia deste trabalho pode ser dividida em duas fases: coleta dos dados e análise dos dados coletados. A primeira atividade para coletar os dados trata-se da criação de scripts para coleta de tweets.

Para coletar tweets foi criado um script que utiliza a API oficial do Twitter. Por meio desta biblioteca é possível coletar cerca de 1% de todos os novos tweets em tempo real, respeitando o limite da API que restringe 50 requisições no intervalo de 15 minutos. Neste trabalho as requisições à API filtraram tweets publicados no Brasil que falam sobre os candidatos à presidência da república, buscando por palavras chaves relacionadas a cada candidato relacionado na Tabela 1.

Os scripts foram desenvolvidos na linguagem Python 3[5] e os dados armazenados em formato CSV (comma-separated-values). Com isso, foram coletados uma amostra de 2 mil tweets para cada candidato totalizando 8 mil tweets, os tweets estão distribuídos em coletas feitas nos dias 14/09/2022, 24/09/2022, 27/09/2022, 28/09/2022 e 02/10/2022.

Na segunda fase deste trabalho, foi feita a análise dos conteúdos coletados. Como os dados são textos em linguagem natural e como será uma grande quantidade de tweets, torna-se inviável a inspeção manual da base de dados. Portanto, foi utilizado técnicas de processamento de linguagem natural para analisar os conteúdos coletados. Após a coleta, todos os tweets foram classificados com um sentimento relacionado ao seu conteúdo podendo ser (positivo ou negativo), essa etapa foi feita usando algoritmos de aprendizagem de máquina supervisionado usando a biblioteca sklearn[6].

As pesquisas de intenção de voto que foram usadas neste trabalho, foram coletadas dos sites do Ipespe, Datafolha e Poder Data. Foram coletadas as pesquisas mais recentes de cada instituto de pesquisa mencionado até o dia 18 de setembro de 2022, para ser usado como comparativo com as últimas pesquisas antes da eleição.

Foram escolhidos esses institutos de pesquisa devido a suas expertises nesta área de pesquisa de intenção de votos e pelo fato de serem usados como referência em grandes portais de notícias e em grandes emissoras da televisão aberta (como a Rede Globo de Televisão por exemplo).

Os dados coletados e processados neste trabalho constituem em uma base de dados contendo tweets que fazem menção aos candidatos listados na Tabela 1. Com essa base em mão, foram feitas análises usando a ferramenta Power BI na etapa de Analisar conteúdos coletados, descrita na Fig 1.

[5] https://www.python.org/download/releases/3.0/

[6] https://scikit-learn.org

IV. RESULTADOS

Usamos como base 2 pesquisas de cada instituto de pesquisa (DataFolha, Poder Data e Ipespe) realizadas em datas diferentes. A primeira delas é a que consta na Tabela 2:

Tabela 2: Pesquisas de intenção de voto

Candidatos	Ipespe	DataFolha	Poder Data	Média
Lula	45%	45%	43%	44%
Bolsonaro	35%	33%	37%	35%
Ciro Gomes	7%	8%	8%	8%
Simone Tebet	5%	5%	5%	5%

As pesquisas relacionadas na Tabela 2 foram realizadas nas seguintes datas: a do Ipespe no dia 17/09/2022, DataFolha 15/09/2022 e Poder Data 13/09/2022. A segunda rodada de pesquisas coletadas pode ser vista na Tabela 3:

Tabela 3: Últimas pesquisas de intenção de voto antes das eleições de primeiro turno

Candidatos	Ipespe	DataFolha	Poder Data	Média
Lula	49%	50%	48%	49%
Bolsonaro	35%	36%	38%	36%
Ciro Gomes	8%	6%	6%	7%
Simone Tebet	7%	5%	5%	6%

As pesquisas relacionadas na Tabela 3 foram feitas nas seguintes datas: a do Ipespe no dia 01/10/2022, DataFolha 01/10/2022 e Poder Data 27/09/2022. A Fig. 2 a seguir mostra uma compilação de pesquisas conduzidas pelo instituto DataFolha e exibidos pelo Jornal Nacional da emissora brasileira Rede Globo de Televisão[7]:

[7] https://g1.globo.com/jornal-nacional/noticia/2022/10/01/datafolha-lula-tem-50percent-dos-votos-validos-e-bolsonaro-36percent.ghtml

Fig. 2 - Pesquisa de intenção de voto para presidente segundo instituto Datafolha.

A Fig. 2 reúne os resultados de pesquisas de intenção de voto no período entre 08 de setembro e 01 de outubro de 2022. Nos gráficos, percebe-se que a situação dos candidatos é relativamente estável ao longo desse período, com pouca variação. A partir deles é possível vislumbrar a possibilidade de vitória do candidato Lula ainda no primeiro turno, considerando-se uma margem de erro de 2 pontos percentuais para mais ou para menos, situação verificada após a segunda quinzena de setembro.

As figuras de 3 a 6 mostram os sentimentos dos tweets em que cada um dos principais presidenciáveis foi mencionado, sendo classificados, conforme detalhado na Seção 3, nas polaridades negativa e positiva.
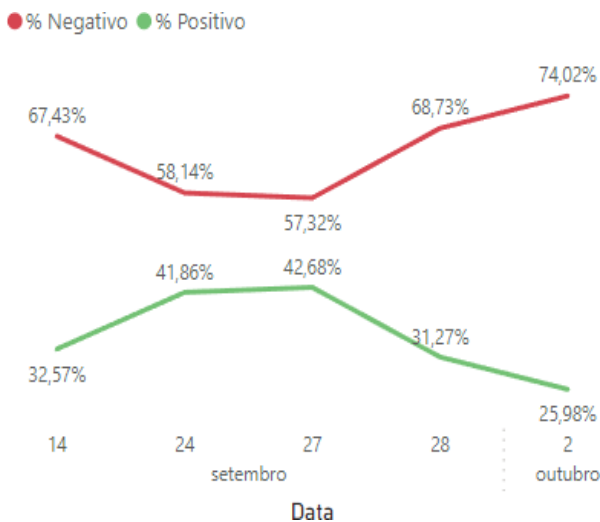
Fig 3 - Sentimentos na rede Twitter - Jair Bolsonaro

O candidato Jair Bolsonaro, teve no dia 14/09/2022[8] um percentual de 32,57% de tweets positivos que chegou a aumentar até 42,86% no dia 27/09/2022, porém no dia 02/10/2022 terminou apenas com 25,98%, É interessante observar o aumento do sentimento negativo em relação a este candidato à medida em que as eleições se aproximaram e a consequente diminuição do sentimento positivo. Uma hipótese para o aumento desse sentimento negativo seria o crescimento do presidente Lula nas pesquisas, com possibilidade do vencimento do candidato petista ainda no primeiro turno. Isso pode ter entusiasmado os eleitores de Lula que aumentaram engajamento no Twitter no sentido de apoiar o petista e expressar um sentimento negativo em relação ao principal adversário, Jair Bolsonaro, na tentativa de desqualificá-lo.
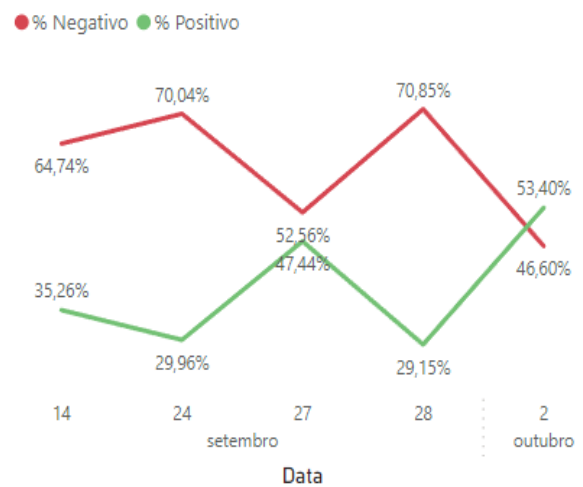
Fig 4 - Sentimentos na rede Twitter – Ciro Gomes

O candidato Ciro Gomes teve bastante oscilação no seu percentual no período de dados coletados, mas teve um aumento significativo do primeiro dia de coleta de 35,26% para 53,40% de tweets positivos no último dia. Nas pesquisas de intenção de voto o candidato em questão teve uma diferença negativa, estando com 8% nas primeiras pesquisas e caindo para 7% na última. Para esse candidato, o sentimento positivo no Twitter (que aumentou na última semana) foi na contramão da sua queda nas pesquisas de intenção de voto, conforme pode-se verificar no contraste entre as Tabelas 1 e 2 também na Fig. 2.

[8] https://g1.globo.com/jornal-nacional/noticia/2022/10/01/datafolha-lula-tem-50percent-dos-votos-validos-e-bolsonaro-36percent.ghtml
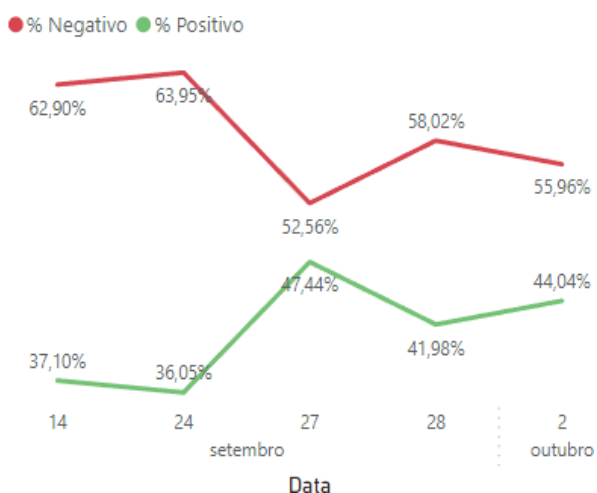
Fig 5 - Sentimentos na rede Twitter – Lula

O candidato Lula teve um aumento de percentual de tweets positivos nos dias de coleta, passando de 37,10% no primeiro dia para 44,04% no último dia. Embora tenha sofrido algumas quedas dentro do período de coleta, houve uma evolução do sentimento positivo em relação ao candidato se comparamos o primeiro e o último dia de coleta dos tweets, o que pode ser um reflexo das pesquisas dos principais institutos apontando-o como favorito e mostrando reais possibilidades de vitória ainda no primeiro turno.
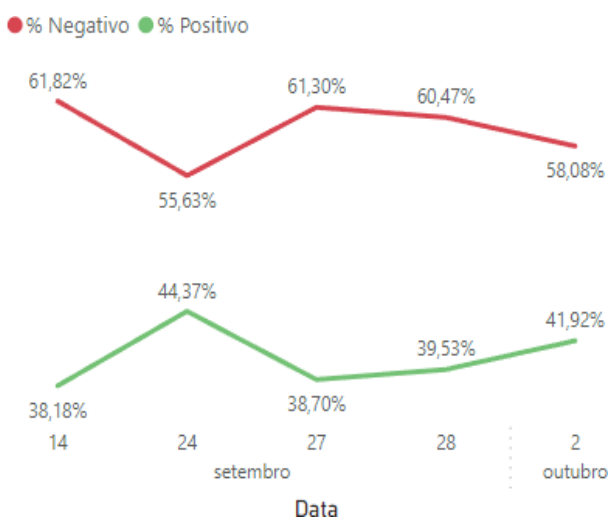
Fig 6 - Sentimentos na rede Twitter – Simone Tebet

A Candidata Simone Tebet, teve um aumento de 38,18% de tweets positivos no primeiro dia de coleta para 41,92% no segundo dia e nas pesquisas também teve um aumento de 5% para 6%. Percebe-se no gráfico uma evolução sutil, porém constante, dos tweets positivos dessa candidata ao longo da semana que antecedeu a eleição. Isso também está em consonância com a sua arrancada na reta final das pesquisas de intenção de votos onde a candidata chegou até mesmo a ultrapassar o candidato Ciro Gomes de acordo com alguns institutos de pesquisa.

Vale ressaltar que o algoritmo de análise de sentimento usado neste trabalho não faz distinção entre perfis fakes (falsos) e perfis autênticos. Sendo assim, não se pode descartar a possibilidade dos resultados serem deturpados em relação à realidade devido ao uso de robôs com o intuito de engajar artificialmente algum candidato.

V. CONCLUSÃO

Observou-se, por meio deste estudo, que o sentimento positivo ou negativo na rede social Twitter oscilou em consonância aos resultados das pesquisas de intenção de voto, com exceção do candidato Ciro Gomes. Todavia, essa oscilação nem sempre se dá na mesma magnitude dos resultados da pesquisa. Conforme verificado na Figura 2 exibida na Seção Estudos Realizados, os candidatos se mantiveram relativamente estáveis na pesquisa de intenção de voto do instituto DataFolha ao longo do mês de setembro, enquanto a evolução dos sentimentos na rede se deu de forma mais intensa. A possibilidade de vitória do candidato petista ainda no primeiro turno pode ter contribuído consideravelmente para o aumento da rejeição de Bolsonaro e diminuição da rejeição de Lula na última semana antes das eleições de primeiro turno. Podemos levantar a hipótese de que a comunidade artística também tenha contribuído para um maior engajamento de Lula no Twitter (artistas brasileiros como Anitta ou Youtubers brasileiros como Felipe Neto). Note que isso é apenas uma hipótese e carece de uma análise mais aprofundada, embora não possa ser descartada como possibilidade e seja interessante que se investigue essa questão em um eventual trabalho futuro.

O Twitter tem um grande potencial como plataforma de exposição de opinião, mas vale ressaltar que os resultados revelados nesta pesquisa sobre o sentimento das postagens não substituem as pesquisas dos institutos tradicionais, já que estes fazem uso de diversos procedimentos metodológicos e estatísticos.

Percebe-se também, se considerarmos principalmente o candidato Jair Messias Bolsonaro, divergências entre os resultados dos institutos de pesquisa, a classificação obtida com os Tweets e o resultado das eleições. Apesar de seguirem uma metodologia criteriosa, não estão isentos de críticas. Não cabe a esse estudo avaliar por qual motivo os principais institutos de pesquisa tem falhado, mas podemos aventar algumas hipóteses que poderão ser discutidas em trabalhos futuros: Os institutos de pesquisa utilizam o censo do Instituto Brasileiro de Geografia e Estatística (IBGE) para nortearem a delimitação de uma amostra populacional que permita captar a heterogeneidade da

população brasileira. De acordo com o site Censo Demográfico 2022, o censo mais recente realizado é de 2022[9]. Contudo, os dados mais recentes disponíveis (publicados) são o de 2010. Como a referência para os institutos de pesquisa têm sido os dados de 2010, a pesquisa pode ter sido prejudicada. Outra hipótese para a discrepância são a dos votos ocultos: pessoas entrevistadas pelos pesquisadores poderiam não estar declarando seu voto. Isso exigiria uma mudança de metodologia. Há também de se levar em consideração que a entrevista pessoal é mais confiável que a por telefone mas tem um problema de timing, já que é mais difícil fazer entrevista pessoalmente na véspera da eleição. Ressalta-se também que na entrevista por telefone, adotada em muitas pesquisas, há menor possibilidade de alcançar os mais pobres, o que poderia estar contribuindo para enviesá-la. Há também a hipótese do voto envergonhado: muitos não têm coragem de dizer que vão votar em candidato A ou B e então declaram-se como indecisos ou que votarão branco ou nulo. Todas essas hipóteses precisam ser avaliadas e isso foge ao escopo deste trabalho.

Também é importante ressaltar o sentimento negativo dos Tweets, claramente predominante em todos os candidatos avaliados, o que nos remete a uma insatisfação generalizada da população que utiliza a rede Twitter com os principais presidenciáveis e com a política de maneira geral. Isso é compatível com a alta rejeição aferida pelos institutos de pesquisa como o Datafolha[10].

Por fim, há evidências de que a rede social Twitter pode refletir, embora geralmente com uma magnitude maior, o sentimento de animação ou desânimo da população em relação às pesquisas. Isso revela a força desta ferramenta enquanto instrumento de divulgação política. Contudo, ressalta-se que não seria apropriado utilizá-la em substituição às pesquisas tradicionais pois há o risco de enviesar a pesquisa não contemplando, por exemplo, pessoas humildes que não dispõe de internet ou que não utilizam redes sociais. Como trabalhos futuros, pretende-se conduzir novas rodadas de pesquisa nas próximas eleições presidenciais, abarcando um período maior com o propósito de verificar se o sentimento das redes têm se aproximado ainda mais dos resultados das pesquisas dos institutos tradicionais.

AGRADECIMENTOS

Ao Instituto Federal de Educação, Ciência e Tecnologia do Norte de Minas Gerais (IFNMG) – Campus Januária, instituição onde sou discente do curso de Bacharelado em Sistemas de Informação, pelo apoio concedido por meio da disponibilização de infraestrutura (laboratório e computadores) para realização dessa pesquisa, tal como o apoio do meu professor e orientador, Raphael Magalhães Hoed.

REFERÊNCIAS

[1] AMARAL, Marcelo Santos; PINHO, José Antonio Gomes. Eleições Parlamentares no Brasil: O Uso do Twitter na Busca por Votos, 2018, DOI: doi.org/10.1590/1982-7849rac2018170269. Disponível em: https://www.scielo.br/j/rac/a/R9fZC87tjCyQJ8hb7jKpzmz/?lang=pt&stop=previous&format=html#. Acesso em 8 out. 2022.

[2] Araújo, M., Gonçalves, P., & Benevenuto, F. (2013). Measuring sentiments in online social networks. 19th Brazilian Symposium on Multimedia and the Web - WebMedia '13. https://doi.org/10.1145/2526188.2526196

[3] BOYD, Danah M.; ELLISON, Nicole B. Social network sites: Definition, history, and scholarship. Journal of computer-mediated Communication, v. 13, n. 1, p. 210-230, 2007.

[4] CAETANO, Josemar Alves et al. Utilizando análise de sentimentos para definição da homofilia política dos usuários do twitter durante a eleição presidencial americana de 2016. In: Congresso da Sociedade Brasileira de Computação-CSBC. 2017.

[5] DA COSTA CÔRTES, Sérgio; PORCARO, Rosa Maria; LIFSCHITZ, Sérgio. Mineração de dados-funcionalidades, técnicas e abordagem PUC, 2002.

[6] GOMES CASTANHEIRA, Luciana. Aplicação de técnicas de mineração de dados em problemas de classificação de padrões. UFMG, 2008.

[7] GOMES PESSANHA, G. R.; OLIVEIRA FIDELIS, T.; DOURADO FREIRE, C.; ALMEIDA SOARES, E. #FIQUEEMCASA: ANÁLISE DE SENTIMENTO DOS USUÁRIOS DO TWITTER EM RELAÇÃO AO COVID19. HOLOS, [S. l.], v. 5, p. 1–20, 2020. DOI: 10.15628/holos.2020.11147. Disponível em: https://www2.ifrn.edu.br/ojs/index.php/HOLOS/article/view/11147 . Acesso em: 8 out. 2022.

[8] JUNQUEIRA K. T. C; FERNANDES A. M. R, Análise de Sentimento em Redes Sociais no Idioma Português com Base em Mensagens do Twitter. 2018, DOI: https://doi.org/10.14210/cotb.v0n0.p681-690. Disponível em: https://periodicos.univali.br/index.php/acotb/article/view/12828. Acesso em 8 out. 2022.

[9] KANSAON, Daniel P.; BRANDÃO, Michele A.; DE PAULA PINTO, Saulo A. Análise de Algoritmos de Classificação para Detecção de Emoções em Tweets em Português Brasileiro. iSys-Brazilian Journal of Information Systems, v. 12, n. 3, p. 116-138, 2020.

[10] MARTINS, Renato F.; PEREIRA, Adriano; BENEVENUTO, Fabrício. An approach to sentiment

[9] https://censo2022.ibge.gov.br

[10] https://www.dinheirorural.com.br/datafolha-no-segundo-turno-bolsonaro-tem-rejeicao-de-51-46-rejeitam-lula/

analysis of web applications in portuguese. In: Proceedings of the 21st Brazilian Symposium on Multimedia and the Web. 2015. p. 105-112.

[11] VIDAL, Delcia Mattos; SOUZA, Danielle Siqueira. Mulheres no espaço político. Afinidades e discrepância em postagens no Twitter. Universidad ORT Uruguay, 2022.

[12] Weis, Indurkhya e Kaufmann. 1999. Predictive Data Mining A Practical Guide. Morgan Kaufmann Publishers, Inc. San Francisco, California. 1999.